Codebook-Based Quantized MIMO Feedback for Closed-Loop Transmit Precoding

Man-On Pun, Ron Porat, Phil Orlik, Jinyun Zhang, Toshiyuki Kuze

TR2009-090 December 2009

Abstract

Advanced quantization schemes are proposed for closed-loop transmit precoding over correlated multiple-input multiple-output (MIMO) channels in this work. Unlike the conventional schemes that directly quantize the MIMO channel covariance matrix and feed back each quantized matrix element, the proposed schemes quantize the channel covariance matrix by exploiting the common rank-one codebook shared by mobile stations and base stations. Compared to the conventional quantization schemes, the proposed quantization schemes can achieve comparable throughput performance with more than 50% overhead reduction at the cost of affordable increase in computational complexity.

IEEE Asilomar Conference on Signals, Systems, and Computers

CODEBOOK-BASED QUANTIZED MIMO FEEDBACK FOR CLOSED-LOOP TRANSMIT PRECODING

Man-On Pun, Ron Porat, Philip Orlik, Jinyun Zhang and Toshiyuki Kuze

ABSTRACT

Advanced quantization schemes are proposed for closed-loop transmit precoding over correlated multiple-input multiple-output (MIMO) channels in this work. Unlike the conventional schemes that directly quantize the MIMO channel covariance matrix and feed back each quantized matrix element, the proposed schemes quantize the channel covariance matrix by exploiting the common rank-one codebook shared by mobile stations and base stations. Compared to the conventional quantization schemes, the proposed quantization schemes can achieve comparable throughput performance with more than 50% overhead reduction at the cost of affordable increase in computational complexity.

Index Terms — Quantized feedback, closed-loop transmit precoding, MIMO.

1. INTRODUCTION

Closed-loop (CL) transmit precoding for multiple-input multiple output (MIMO) downlink (DL) transmission has been well proven as an effective and practical capacity-achieving technique [1]. Indeed, it has been standardized in many emerging communications systems such as IEEE 802.16m [2] and LTE [3]. In CL transmit precoding, a mobile station (MS) first estimates the DL channel before feeding back the estimated channel state information (CSI) to the base station (BS). Generally speaking, two different approaches are commonly employed in returning the estimated CSI to the BS.

Assuming that the MS and BS share a common codebook, the first approach is designed to let the MS first derive the optimal precoding codeword from the common codebook by exploiting the estimated instantaneous CSI. After that, the MS feeds back the optimal codeword index to the BS. Upon receiving the feedback, the BS precodes the DL data with the designated codeword in the subsequent DL transmission until it receives the next precoding codeword from the MS. This approach has been demonstrated very effective in improving system throughput over *uncorrelated* MIMO channels.

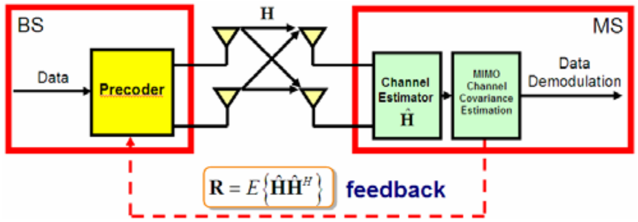

Fig. 1 System schematic of closed-loop transmit precoding based on long-term channel covariance matrix

However, the first approach is handicapped by *correlated* MIMO channels due to the fact that existing codebooks are unitary and designed for uncorrelated MIMO channels. To circumvent this obstacle, a second approach has been recently proposed to feed back the long-term channel covariance matrix to the BS as illustrated in Fig. 1 [4]. Exploiting the long-term channel covariance matrix, both BS and MS adaptively employ transformed codebooks to optimize the DL precoding. In addition to the performance improvement, the second approach requires less frequent feedback due to the nature of the feedback as compared to the first approach. Despite the aforementioned advantages, unlike the first approach in which the feedback information contains already quantized codeword indices, the MIMO channel covariance matrix has to be first quantized before it can be practically fed back to the BS.

The conventional quantization method performs direct quantization on the covariance matrix [2]. More specifically, it employs different levels of quantization precision on the diagonal and upper-triangular elements by exploiting the Hermitian structure of the covariance matrix. Unfortunately, this simple scheme incurs large feedback overhead.

In this work, we propose reduced-feedback quantization schemes by exploiting the common codebooks shared by the BS and MS. Rather than directly quantizing the MIMO channel covariance matrix, quantization schemes are developed to represent the matrix by using rank-one codewords either contained in the common codebooks or in a pre-defined steering vector form. As a result, the proposed

---

Man-On Pun, Philip Orlik and Jinyun Zhang are with Mitsubishi Electric Research Laboratories, 201 Broadway Cambridge, MA 02139, U.S.A. (Emails: {mpun, porlik, jzhang}@merl.com)

Ron Porat, is with InterDigital, 781 Third Avenue, King of Prussia, PA 19406, U.S.A. (Email: Ron.Porat@InterDigital.com)

Toshiyuki Kuze is with Mitsubishi Electric Corporation, 5-1-1 Ofuna Kamakura, Kanagawa 2478501, Japan.
(Email: Kuze.Toshiyuki@ah.MitsubishiElectric.co.jp)

covariance matrix quantization schemes can achieve good quantization accuracy with substantially reduced feedback overhead at the cost of some affordable increase in computational complexity. It should be emphasized that the proposed quantized feedback schemes can be also employed for feeding back the interference covariance matrix for applications such as interference nulling and multi-user MIMO (MU-MIMO).

Notation: Vectors and matrices are denoted by boldface letters. Furthermore, we use $E\{\cdot\}$, $(\cdot)^*$ and $(\cdot)^H$ for expectation, conjugate and Hermitian transposition, respectively.

2. PROBLEM FORMULATION

We denote by $\mathbf{R}$ the DL MIMO channel covariance matrix. It should be emphasized that $\mathbf{R}$ represents the channel covariance matrix over a particular subcarrier or resource block in orthogonal frequency division modulation (OFDM)-based systems. Exploiting the fact that $\mathbf{R}$ is Hermitian with the real-valued diagonal elements much larger than the absolute values of the complex-valued off-diagonal elements, a simple direct quantization method has been proposed in IEEE 802.16m. More specifically, each diagonal element is quantized with one bit while each upper off-diagonal element with four bits, as shown in Fig. 2. As a result, for $N_t = 4$, it requires total 28 bits or four bytes to quantize $\mathbf{R}$. In the sequel, this scheme is referred to as the direct quantization method (DQM).

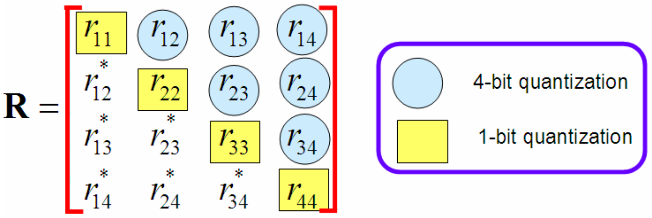

Fig. 2 Conventional direct quantization method proposed in IEEE 802.16m for for $N_t = 4$ [2].

Clearly, the amount of overhead required by DQM grows in the order of N_t^2 for a matrix of dimension $N_t \times N_t$. Furthermore, for OFDM systems, the total feedback overhead increases proportionally with the number of subcarriers or resource blocks. Thus, it is desirable to reduce the quantized feedback overhead for each $\mathbf{R}$. For presentational simplicity, we concentrate our following discussion on $N_t = 4$ while the discussion can be extended to other values of N_t in a straightforward manner.

Alternatively, we observe that the covariance matrix $\mathbf{R}$ can be decomposed into the following form.

$$\mathbf{R} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H \qquad (1)$$

where $\mathbf{U}$ and $\boldsymbol{\Lambda}$ are 4×4 unitary and diagonal matrices, respectively. Recall that the MS and BS share multiple common unitary codebooks of different ranks. Intuitively speaking, if we can find a codeword $\mathbf{V}$ in the rank-4 common codebook such that $\mathbf{V}^H\mathbf{U} \approx \mathbf{I}$, then only the index of the chosen rank-4 codeword and four quantized real-valued $\{\lambda_i; i = 1,2,3,4\}$ are required to be returned to the BS, which stands for a substantial feedback overhead reduction compared to DQM. However, two problems associated with this approach arise. First, the rank-4 codebook defined in the standards is of a much smaller size compared to the lower-rank codebooks. As a result, it is not guaranteed that we can find a rank-4 codeword satisfying the requirement of $\mathbf{V}^H\mathbf{U} \approx \mathbf{I}$. Second, considering a rank-4 codebook of B bits, the computational complexity involved in exhaustively searching the optimal $\mathbf{V}$ is of the order of $O\left(2^B \cdot N_t^3\right)$, which can be prohibitively expensive for a large N_t. In particular, the second problem becomes more challenging for OFDM systems with a large number subcarriers or resource blocks. In the next section, we propose rank-one codebook-based quantization schemes to cope with these two problems.

3. PROPOSED SCHEMES

We first rewrite (1) into the following form.

$$\mathbf{R} = \sum_{p=1}^{P} \lambda_p \cdot \mathbf{u} \cdot \mathbf{u}^H , \qquad (2)$$

where $\mathbf{u}_p$ and λ_p are the eigenvectors of length $N_t \times 1$ and real-valued eigenvalues of $\mathbf{R}$ with $\operatorname{rank}(\mathbf{R}) = P \le N_t$. Without loss of generality, we assume $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_P$. For correlated channels under consideration, we have $P < N_t$. Thus, for $N_t = 4$, we can safely assume that $P = 1$ or $P = 2$ in practice.

Next we assume that the MS and BS share a common rank-one codebook of B bits, denoted by $\{\mathbf{v}_b; b = 1,2,\cdots,2^B\}$. Thus, if we can find two rank-one codewords in the rank-one codebook, $\{\mathbf{v}_1, \mathbf{v}_2\}$, that match the first two principal vectors $\{\mathbf{u}_1, \mathbf{u}_2\}$, then only information about the two codeword indices and the ratio between their corresponding eigenvalues is sufficient for the BS to reconstruct $\mathbf{R}$ up to an unknown scalar. More specifically, if the common codebook is comprised of 2^B rank-one codewords and the ratio between the two eigenvalues is quantized with ℓ bits, the total number of feedback bits is given by $2B + \ell$.

To reduce computational complexity, we develop sub-optimal search algorithms to find the best-matching rank-one codewords sequentially, rather than jointly. In the following, we will first propose a Gram-Schmidt-based searching algorithm (GSSA) before exploring different variations of the algorithm.

The GSSA is summarized in the following steps.

Step 1: For a given rank-one codebook $\left\{\mathbf{v}_b\right\}$, we first find the codeword that best matches the first principal eigenvector as

$$m = \underset{b=1,2,\cdots,2^B}{\arg\max}\, \mathbf{v}_b^H \mathbf{R} \mathbf{v}_b \, , \tag{3}$$

with $\lambda_1 = \mathbf{v}_m^H \mathbf{R} \mathbf{v}_m$.

Step 2: Next, we generate a new set of rank-one vectors $\left\{\mathbf{d}_b\right\}$ that are orthogonal to $\mathbf{v}_m$ by using the Gram-Schmidt method:

$$\mathbf{d}_b = \frac{\mathbf{v}_b - \left(\mathbf{v}_m^H \mathbf{v}_b\right)\mathbf{v}_m}{\left\|\mathbf{v}_b - \left(\mathbf{v}_m^H \mathbf{v}_b\right)\mathbf{v}_m\right\|} \, , \tag{4}$$

for $b = 1,2,\cdots,2^B$.

Step 3: After that, we find a codeword in $\left\{\mathbf{d}_b\right\}$ that best matches the second principal eigenvector of as

$$n = \underset{b=1,2,\cdots,2^B}{\arg\max}\, \mathbf{d}_b^H \mathbf{R} \mathbf{d}_b \, , \tag{5}$$

with $\lambda_2 = \mathbf{d}_n^H \mathbf{R} \mathbf{d}_n$.

Step 4: Finally, we quantize the ratio $q = \lambda_2 / \lambda_1$ into a predefined set of thresholds.

The feedback includes the two indices of the winning eigenvectors $\left\{m, n\right\}$ and the index of the *quantized* eigenvalue ratio q .

It is worthwhile to point out several interesting properties related to the proposed rank-one quantization approach. First of all, because the rank-one common codebook is usually much larger than the higher-rank common codebooks, it is more likely for the proposed approach to find a good matching codeword for a given principle eigenvector. Furthermore, the total computational complexity of GSSA is about $O\left(2^{B+1} \cdot N_t^2\right)$ with most computation incurred in (3) and (5). This stands for a significant computational reduction compared to (1) of $O\left(2^B \cdot N_t^3\right)$.

Based on the above Gram-Schmidt algorithm, we can also derive algorithms with different variations. For instance, rather than using the optimization objective function in Step 3, we can employ the following optimization function.

$$n = \underset{b=1,2,\cdots,2^B}{\arg\min}\, \| \mathbf{R} - \lambda_1 (\mathbf{v}_m \mathbf{v}_m^H + q \mathbf{d}_b \mathbf{d}_b^H) \|^2 . \tag{6}$$

4. SIMULATION RESULTS

In this section, system-level simulation results are provided to confirm the performance of the rank-one quantized feedback schemes proposed in the previous section. We compare the performance of different quantized feedback schemes in terms of the DL spectral efficiency, assuming a minimum mean squared error (MMSE) receiver at the MS with perfectly known interference. We use the 802.16m 4-antenna codebooks (6bit and 4bit subset) and also examine the performance of a 4-bit steering vector given below to quantize the strongly corrected channels:

$$\mathbf{sv}_i = \frac{1}{2}\begin{bmatrix} 1 \\ \mathrm{e}^{j\pi\sin(\theta_i)} \\ \mathrm{e}^{j2\pi\sin(\theta_i)} \\ \mathrm{e}^{j3\pi\sin(\theta_i)} \end{bmatrix} , \tag{7}$$

where $\theta_i = (i - \frac{1}{2})\frac{\pi}{24} - \frac{\pi}{3}$ with $i = 1,...,16$.

As shown in the simulation, the steering-vector approach is particularly helpful in reducing computation and feedback overhead for correlated antennas.

We examine the performance of the following six quantized feedback schemes:

1. ‘**Perfect SV**’ uses the strongest unquantized singular vector of the average narrowband transmit correlation matrix. The total feedback overhead is infinite;
2. ‘**4bit+R**’ uses the four-bit codebook subset transformed by the unquantized $\mathbf{R}$. The total feedback overhead is infinite;
3. ‘**4bit+VQ (6+6+1)**’ uses the proposed vector quantization scheme with the 6-bit rank-one codebook specified in the IEEE 802.16m. The total feedback overhead is 13 bits;
4. ‘**4bit+(4+4+1)**’ uses the proposed 4-bit steering-vector quantization scheme as shown in (7). The total feedback overhead is 9 bits;
5. ‘**4bit+Quantized R**’ uses the conventional scheme specified in the IEEE 802.16m to perform element-wise quantization. The total feedback overhead is 28 bits;
6. ‘**6 bit**’ uses the 802.16m rank-1 6 bits codebook only (no feedback of $\mathbf{R}$).

Clearly, ‘Perfect SV’ and ‘4bit+R’ stand for the ideal feedback cases that cannot be realized in practice.

In the simulation, the Macro and Urban Macro spatial channel models (SCM) with an angular spread of 15^o are employed [2]. Two types of antennas configurations are simulated, namely the ULA with four ${}^{\lambda}\!/_{2}$-spaced antennas and two ${}^{\lambda}\!/_{2}$-spaced cross-polarized array. We simulate an

OFDM system similar to that in the IEEE 802.16m standard [2]. We set the DL bandwidth to four physical resource blocks (equivalent to 800KHz), assuming one precoder per band. To model the fading environment and system feedback delay, we set the MS’ mobile speed to 3kmph while the delay equal to 5ms. Furthermore, we assume that the DL channel estimation is perfect. Finally, up to four users are selected from a 4-user pool by exhaustive selection such that the total throughput is maximized.

To investigate the performance of the proposed schemes, we adopt the IEEE 802.16m rank-one codebook in our simulation. The ratio of the second to first eigenvalues is quantized to the nearest point to either 0.25 or 0.5, which requires only one-bit feedback.

The plots show the spectral efficiency ratio of a system employing various feedback schemes as compared to a feedback of the 4bit codebook subset only (without R).

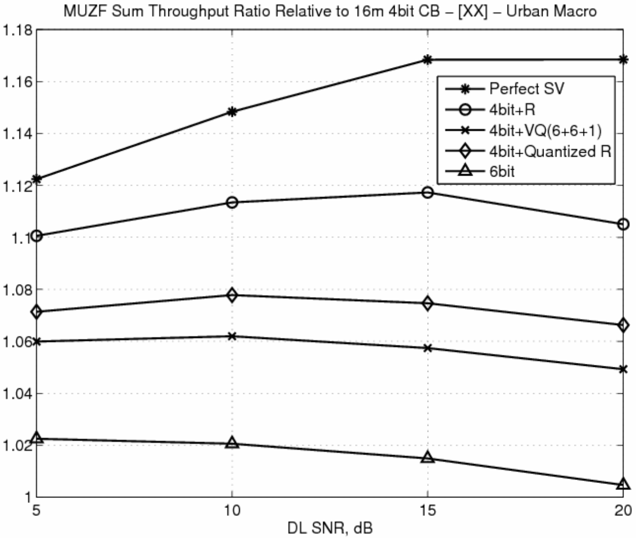

Fig. 3 DL spectral efficiency ratio as a function of SNR in Urban Macro SCM channel with cross polarized antennas.

Fig. 3 shows the DL spectral efficiency ratio as a function of SNR in Urban Marco SCM channel with cross polarized antennas. Inspection of Fig. 3 suggests that the proposed ‘4bit+VQ (6+6+1)’ scheme achieves comparable performance as the conventional DQM ‘4 bit + Quantized R’ with less than 50% feedback overhead. Compared to the ideal ‘Perfect SV’, the proposed scheme has about 20% loss of DL spectral efficiency.

Fig. 4 depicts the DL spectral efficiency in the Urban Marco SCM channel with ULA. We can observe the similar performance trend as shown in Fig. 3.

Finally, we examine the performance of the quantized feedback schemes in the Suburban Marco SCM channel with ULA. Interestingly, Fig. 5 indicates that the two proposed quantized feedback schemes, namely, ‘4bit+VQ (6+6+1)’ and ‘4bit+VQ (4+4+1)’, outperform the conventional DQM ‘4bit+Quantized R’ by about 5% with 50% less feedback overhead.

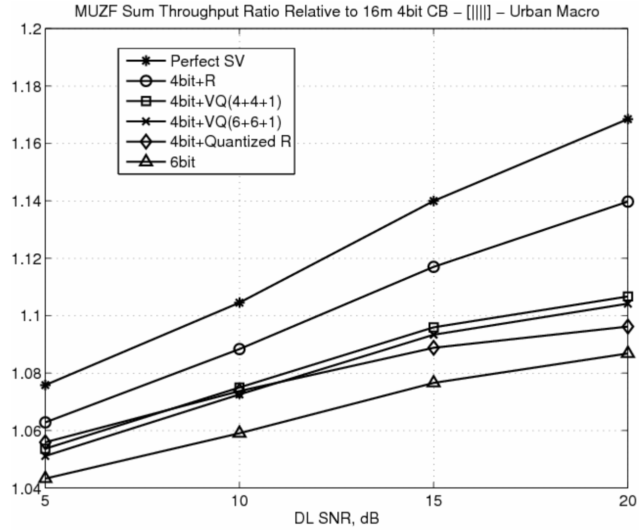

Fig. 4 DL spectral efficiency ratio as a function of SNR in Urban Macro SCM channel with uniform linear array.

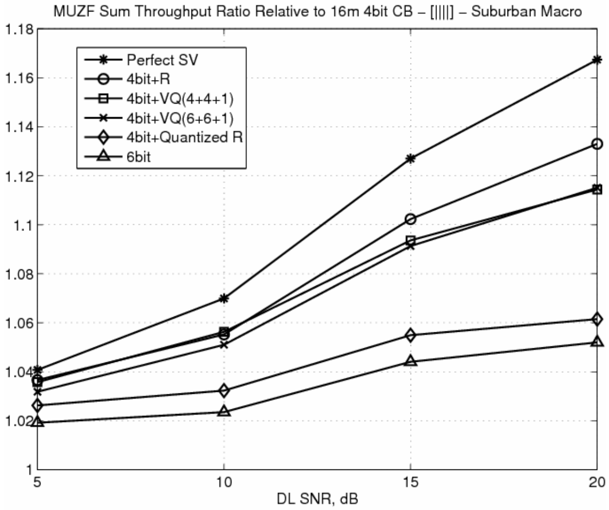

Fig. 5. DL spectral efficiency ratio as a function of SNR in Suburban Macro SCM channel with uniform linear array.

5. APPLICATION TO BEAMFORMING AND NULLING

Beamforming and nulling is a well known technique for reducing interference to adjacent cell co-channel users while at the same time improving the SINR of the served user. This is more easily facilitated in TDD systems using sounding signals whereby each BS estimates its own user channel and the interference correlation matrix and forms a beamforming solution which is a tradeoff between gain delivered to its user and leakage energy delivered to adjacent cell co-channel users.

The solution typically adopted is $V = P\{(\sum_i R_i + \alpha N_o I)^{-1} R_S\}$ where the operator P{} denotes the principal eigenvector, R_i is the spatial correlation of other-cell users exchanged between the BSs, R_S is the spatial correlation of the serving BS user, α is a regularization factor, N_o is an estimate of the user's total noise plus interfering power and I is an identity matrix.

The same technique can be facilitated in FDD systems by explicitly estimating the interference correlation matrix, quantizing it using the techniques described before and feeding it back to the BS.

Simulation results of various quantization techniques are beyond the scope of this paper.

6. CONCLUSION

Quantized feedback schemes have been devised for closed-loop transmit precoding over correlated MIMO channels by exploiting the common rank-one codebooks shared by the BS and MS or predefined steering-vector codebooks. Simulation results have confirmed that the proposed schemes can achieve comparable throughput performance with more than 50% less feedback overhead, compared to the direct quantization method in the current 16m specification.

REFERENCES

[1] H. Sampath, P. Stoica and A. Paulraj, “Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion,” *IEEE Trans. Comm.*, Vol. 49, pp. 2198-2206, December 2001.

[2] S. Hamiti, “The Draft IEEE 802.16m System Description Document”, IEEE 802.16m-08/003r6.

[3] 3GPP TR 25.913 V7.2.0, “Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN)”, 2006.

[4] V. Raghavan, R.W. Heath and M. Sayeed, “Systematic Codebook Designs for Quantized Beamforming in Correlated MIMO Channels”, *IEEE Journal on Selected Areas in Communications*, Vol. 25, pp. 1298-1310, September 2007.